Sparse Signal Recovery with Multiple Prior Information: Algorithm and Measurement Bounds

Luong, Huynh Van; Deligiannis, Nikos; Seiler, Jürgen; Forchhammer, Søren; Kaup, André

Published in:
Signal Processing

Link to article, DOI:
10.1016/j.sigpro.2018.06.019

Publication date:
2018

Document Version
Peer reviewed version

Citation (APA):
Highlights

- Recovery with multiple prior information via solving n-1 minimization is proposed
- Theoretical measurement bounds required by the n-1 minimization are established
- The derived bounds of the n-1 minimization are sharper
- The proposed n-1 minimization outperforms the state-of-the-art algorithms
Sparse Signal Recovery with Multiple Prior Information: Algorithm and Measurement Bounds

Huynh Van Luong\textsuperscript{a,b,}\textsuperscript{∗}, Nikos Deligiannis\textsuperscript{b}, Jürgen Seiler\textsuperscript{a}, Søren Forchhammer\textsuperscript{c}, André Kaup\textsuperscript{a}

\textsuperscript{a}Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, 91058 Erlangen, Germany
\textsuperscript{b}Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, and imec, B3001 Leuven, Belgium
\textsuperscript{c}DTU Fotonik, Technical University of Denmark, 2800 Lyngby, Denmark

Abstract

We address the problem of reconstructing a sparse signal from compressive measurements with the aid of multiple known correlated signals. We propose a reconstruction algorithm with multiple side information signals (RAMSI), which solves an \( n\ell_1 \) minimization problem by weighting adaptively the multiple side information signals at every iteration. In addition, we establish theoretical bounds on the number of measurements required to guarantee successful reconstruction of the sparse signal via weighted \( n\ell_1 \) minimization. The analysis of the derived bounds reveals that weighted \( n\ell_1 \) minimization can achieve sharper bounds and significant performance improvements compared to classical compressed sensing (CS). We evaluate experimentally the proposed RAMSI algorithm and the established bounds using numerical sparse signals. The results show that the proposed algorithm outperforms state-of-the-art algorithms—including classical CS, \( \ell_1\ell_1 \) minimization, Modified-CS, regularized Modified-CS, and weighted \( \ell_1 \) minimization—in terms of both the theoretical bounds and the practical performance.

Keywords: Compressed sensing, prior information, weighted \( n\ell_1 \) minimization, measurement bounds

1. Introduction

Compressed sensing (CS) [1–15] states that sparse signals can be recovered in a computationally tractable manner from a limited set of measurements by minimizing the \( \ell_1 \)-norm. The CS performance can be improved by replacing the \( \ell_1 \)-norm with a weighted \( \ell_1 \)-norm [8, 9, 16–18]. The studies in [11, 12] provide bounds on the number of measurements required for successful signal recovery based on convex

\textsuperscript{∗}This work was supported by the Humboldt-Fellowship, the Alexander von Humboldt Foundation, Germany. A part of this work, the RAMSI algorithm, has been presented at the 2016 Data Compression Conference [47].

\textsuperscript{∗}Corresponding author

Email addresses: hvanluong@etrovub.be (Huynh Van Luong), ndeligia@etrovub.be (Nikos Deligiannis), juergen.seiler@fau.de (Jürgen Seiler), sofo@fotonik.dtu.dk (Søren Forchhammer), andre.kaup@fau.de (André Kaup)
optimization. Furthermore, distributed compressed sensing [13, 14] allows a correlated ensemble of sparse signals to be jointly recovered by exploiting the intra- and inter-signal correlations.

We consider the problem of reconstructing a signal given side or prior information, gleaned from a set of known correlated signals. Initially, this problem was studied in [16, 19–28], where the modified-CS method [19, 21] considered that a part of the support is available from prior knowledge and tried to find the signal that satisfies the measurement constraint and is sparsest outside the known support. Prior information on the sparsity pattern of the data was also considered in [23] and information-theoretic guarantees were presented. The studies in [24, 25] introduced weights into the $\ell_1$ minimization framework that depend on the partitioning of the source signal into two sets, with the entries of each set having a specific probability of being nonzero.

Alternatively, the studies in [29–32] incorporated side information on CS by means of $\ell_1$-$\ell_1$ minimization and derived bounds on the number of Gaussian measurements required to guarantee perfect signal recovery. It was shown that $\ell_1$-$\ell_1$ minimization can dramatically improve the reconstruction performance over CS subject to a good-quality side information [29, 30]. The study in [33] proposed a weighted-$\ell_1$ minimization method to incorporate prior information—in the form of inaccurate support estimates—into CS. The work also provided bounds on the number of Gaussian measurements for successful recovery when sufficient support information is available. Furthermore, recent studies proposed CS with side information in practical applications, i.e., compressive video foreground extraction [34–36], magnetic resonance imaging (MRI) [17], and synthetic aperture radar imaging [37].

The problem of sparse signal recovery with prior information also emerges in the context of reconstructing a sequence of time-varying sparse signals from low-dimensional measurements [23, 38–42]. The study in [42] reviewed a class of recursive algorithms for recovering a sequence of time-varying sparse signals from a small number of measurements. The problem of estimating a time-varying, sparse signal from streaming measurements was studied in [38, 42], while the work in [23] addressed the recovery problem in the context of multiple measurement vectors. The problem also appears in the context of robust PCA and online robust PCA [39–41], a framework that finds important application in video background subtraction. The studies in [39–41] used the modified-CS [21, 22] method to leverage prior knowledge under the condition of slow support and signal value changes.

1.1. Motivation

Emerging applications require reconstructing a sequence of signals from compressive measurements; for example, dynamic MRI [43] involves the acquisition of moving cardiovascular, abdominal, or larynx images. Related research questions are also encountered in the context of compressive video sensing [44, 45]. Alternative applications [46–48] follow a distributed sensing scenario, where a plethora of tiny heterogeneous devices collect information from the environment, and require sensing and processing.
under time and resource constraints. These challenges can be addressed by leveraging the distributed sparse representation of the multiple sources [48–50]. The problem in this setup is to reconstruct the sparse sources along with exploiting the correlation among them; alias, a key question is how to reconstruct a signal from a small number of measurements, by leveraging efficiently the redundancy gleaned from multiple correlated signals.

Existing attempts to incorporate prior information in compressed sensing are typically considering one known prior information signal of good quality; see for example [17, 29, 30, 34, 35]. Conversely, we are aiming at reconstructing a sparse signal from a small set of measurements with the aid of multiple prior signals, which change in time; that is, there are arbitrary prior information qualities and varying correlations among them. This raises key interesting questions:

- How can we leverage efficiently the correlation across multiple prior information signals? This implies a strategy to exploit the useful information from the multiple signals and alleviate negative effects due to low quality prior information.
- How many measurements are required to successfully reconstruct the sparse signal given multiple prior signals? This calls for bounds on the number of measurements required to guarantee successful signal recovery.

1.2. Contributions

To address these questions, we contribute in a twofold way. Firstly, we propose a novel sparse signal Reconstruction Algorithm that leverages Multiple Side Information (RAMSI). Secondly, we establish lower bounds on the number of measurements required by RAMSI to recover the sparse signal.

The RAMSI algorithm solves a \((re-)weighted n-\ell_1\) minimization problem: Per iteration of the recovery process, the algorithm weights adaptively the \(n\) prior information signals. As such, unlike existing works [17, 29, 30, 34, 35], which exploit only one prior information signal, RAMSI can efficiently leverage the correlations among multiple signals and adapt on-the-fly to changes of the correlations. We show experimentally that RAMSI leads to higher recovery performance than state-of-the-art methods [7, 21, 22, 25, 29].

We also establish measurement bounds for the \(weighted n-\ell_1\) minimization problem, which serve as lower bounds for the RAMSI algorithm. The bounds depend on the support of the source signal to be recovered and the correlations between the target signal and the multiple prior information signals. The correlations are expressed via the supports of the differences between the source and prior information signals. We will show that the \(weighted n-\ell_1\) minimization bounds are sharper compared to those of the classical CS [1, 3, 5] and the \(\ell_1-\ell_1\) reconstruction [29, 30] methods. These bounds depict the advantage of RAMSI to deal with heterogeneous side information signals including possible...
poor prior information signals. Furthermore, we show—both theoretically and practically—that the
performance of the method is improved with the number of available prior information signals.

1.3. Outline

The rest of this paper is as follows: Section 2 reviews the background on CS and CS with side
information. The RAMSI algorithm is proposed in Section 3, whereas our measurement bounds and
their analysis are presented in Section 4. The bounds and the performance of RAMSI are assessed in
Section 5 and Section 6 concludes the work.

2. Background

2.1. Compressed Sensing

Let \( x \in \mathbb{R}^n \) be a signal that is sparse in a domain or learned set of basis. The signal can be reduced
to a vector \( y = \Phi x \in \mathbb{R}^m \) by sampling with a measurement matrix \( \Phi \in \mathbb{R}^{m \times n} \) whose
elements are sampled from an i.i.d. Gaussian distribution. The signal can be recovered by solving the
Basis Pursuit problem [3, 5, 52]:

\[
\min_x \|x\|_1 \ \text{subject to} \quad y = \Phi x, \tag{1}
\]

where \( \|x\|_1 := \sum_{i=1}^n |x_i| \) is the \( \ell_1 \)-norm of \( x \) wherein \( x_i \) is an element of \( x \). Problem (1) becomes an
instance of finding a general solution:

\[
\min_x \{ H(x) = f(x) + g(x) \}, \tag{2}
\]

where \( f := \mathbb{R}^n \to \mathbb{R} \) is a smooth convex function and \( g := \mathbb{R}^n \to \mathbb{R} \) is a continuous convex function,
possibly non-smooth. Specifically, in Problem (1) we have \( g(x) = \lambda \|x\|_1 \) and \( f(x) = \frac{1}{2} \|\Phi x - y\|_2^2 \),
where \( \| \cdot \|_2 \) denotes \( \ell_2 \)-norm, with Lipschitz constant \( L_{\nabla f} [7] \). Using proximal gradient methods [7],
\( x^{(k)} \) at iteration \( k \) is computed as

\[
x^{(k)} = \Gamma_{\frac{1}{L}g} \left( x^{(k-1)} - \frac{1}{L} \nabla f(x^{(k-1)}) \right), \tag{3}
\]

where \( L \geq L_{\nabla f} \), and \( \Gamma_{\frac{1}{L}g} (x) \) is the proximal operator:

\[
\Gamma_{\frac{1}{L}g} (x) = \arg \min_{v \in \mathbb{R}^n} \left\{ \frac{1}{L} g(v) + \frac{1}{2} \|v - x\|_2^2 \right\}. \tag{4}
\]

The classical \( \ell_1 \) minimization problem in CS [1, 3, 5] requires \( m_{\ell_1} \) measurements [11, 29, 30] for
successful reconstruction, bounded as

\[
m_{\ell_1} \geq 2s_0 \log \frac{n}{s_0} + \frac{7}{5}s_0 + 1, \tag{5}
\]

where \( s_0 := \|x\|_0 = |\{ i : x_i \neq 0 \}| \) denotes the number of nonzero elements in \( x \) as the support of \( x \),
with \( |.| \) denoting the cardinality of a set and \( \| \cdot \|_0 \) being the \( \ell_0 \)-pseudo-norm.
2.2. CS with Support Knowledge

The modified-CS method in [19, 21] considered prior knowledge in the form of knowing a part of the signal’s support. Furthermore, the study in [33] proposed a weighted-$\ell_1$ minimization method to incorporate prior support information and provided bounds on the number of measurements—we denote this number by $m_{\omega\ell_1}$—for successful recovery. It is worth noting that the bound of the number of measurements required by modified-CS, $m_{\text{modCS}}$, is a special case of the bound on $m_{\omega\ell_1}$. Namely, without considering weights on the prior support, it holds that [33, Corollary 8]

$$m_{\omega\ell_1} \equiv m_{\text{modCS}} \geq s_0 + (1 + C^{-2}) s_c \log (cn/s_c),$$

where $C < 1$ and $s_c$ is the size of the support estimate error defined by $|(\tilde{T} \cap T^c) \cup (T \cap \tilde{T}^c)| \leq s_c$, where $T$ is the support, $\tilde{T}$ is the support estimate and $T^c$ and $\tilde{T}^c$ their respective complements. In [33], it is assumed that $s_0 \leq n/2$ and $s_c \leq s_0$. Due to $C < 1$, we can write

$$m_{\omega\ell_1} \equiv m_{\text{modCS}} \geq 2s_c \log(n/s_c) + 2s_c + s_0.$$  \hspace{1cm} (6)

2.3. CS with $\ell_1$-$\ell_1$ Minimization

The $\ell_1$-$\ell_1$ minimization approach [29, 30, 34] reconstructs $x$ given a signal $z \in \mathbb{R}^n$ as prior information by solving the following problem:

$$\min_x \left\{ \frac{1}{2} \| \Phi x - y \|_2^2 + \lambda (\| x \|_1 + \| x - z \|_1) \right\}.$$  \hspace{1cm} (7)

The bound on the number of measurements required by Problem (7) to successfully reconstruct $x$ depends on the quality of the side information signal $z$ as [29, 30, 34]

$$m_{\ell_1-\ell_1} \geq 2\bar{h} \log \left( \frac{n}{s_0 + \xi/2} \right) + \frac{7}{5} \left( s_0 + \frac{\xi}{2} \right) + 1,$$  \hspace{1cm} (8)

where

$$\xi := |\{ i : z_i \neq x_i = 0 \}| - |\{ i : z_i = x_i \neq 0 \}|,$$  \hspace{1cm} (9a)

$$\bar{h} := |\{ i : x_i > 0, x_i > z_i \}| \cup \{ i : x_i < 0, x_i < z_i \}|,$$  \hspace{1cm} (9b)

wherein $x_i, z_i$ are corresponding elements of $x, z$. It has been shown that the solution of Problem (7) improves over the solution of Problem (1) provided that the prior information has good enough quality [29, 30]. The quality is expressed by a high number of elements $z_i$ that are equal to $x_i$, thereby leading to $\xi$ in (9a) being small.

3. Recovery With Multiple Prior Information

3.1. Problem Statement

We consider the problem of recovering $x$ from a low-dimensional measurement vector $y$ given prior information. Classical CS methods can be used to recover $x$ from $y$; however, these methods do
not leverage the correlations across multiple different data vectors. Examples of practical applications that motivate our work include compressive video sensing and analysis [35, 51, 53] and dynamic MRI reconstruction [45]. For instance, consider the compressive video sensing setup illustrated in Fig. 1(a): an incoming video frame is sensed using compressive measurements at the encoder and reconstructed at the decoder with the aid of multiple prior information, gleaned from a set of previously reconstructed frames. Fig. 1(b) depicts the vectorized wavelet transform coefficients (a.k.a. sparse representation) of frame no. 1286 in the Bootstrap video sequence—which corresponds to the target signal $x$ in our scenario—and the wavelet representations of the two previous reconstructed frames $z_1$ and $z_2$—which respectively correspond to frame no. 1285 and 1284 in the Bootstrap video sequence. It is clear that the sparse representation $x$ is highly correlated with $z_1$ and $z_2$; this correlation can be leveraged to improve the reconstruction of $x$. Other application scenarios include compressive foreground extraction [35] and compressive online robust principle component analysis [53], where a foreground frame $x$ can be reconstructed from previously reconstructed foregrounds $z_1$ and $z_2$.

---

1Alternatively, the prior information could be gleaned from motion-compensated versions of the reconstructed frames [54].
The ℓ1-ℓ1 minimization framework in Problem (7) can be used to recover x from y given either z1 or z2, that is, only one prior information vector. Moreover, it can be that ℓ1-ℓ1 minimization performs worse than the ℓ1 minimization method due to low correlation between the target signal and the prior information signal, e.g., z2. To address these two limitations, we propose a new reconstruction algorithm with multiple side information (RAMSI). The input of RAMSI is the measurement vector \( y = \Phi x \) and \( J \) prior information signals \( z_1, \ldots, z_J \in \mathbb{R}^n \). The objective function is constructed by using an n-ℓ1-norm function in Problem (2):

\[
g(x) = \lambda \sum_{j=0}^{J} \| W_j(x - z_j) \|_1,
\]

where \( z_0 = 0 \) and \( W_j \) are diagonal weight matrices, \( W_j = \text{diag}(w_{j1}, w_{j2}, \ldots, w_{jn}) \), wherein \( w_{ji} > 0 \) is the weight in \( W_j \) at index \( i \). Namely, the objective function of the proposed n-ℓ1 minimization problem is given by

\[
\min_{x} \left\{ H(x) = \frac{1}{2} \| \Phi x - y \|_2^2 + \lambda \sum_{j=0}^{J} \| W_j(x - z_j) \|_1 \right\}.
\]

3.2. The Proposed RAMSI Algorithm

We solve Problem (11) by iteratively computing \( x \) and updating \( W_j \) based on the proximal gradient method [7]. The question is then how to determine the weight values in order to improve the reconstruction by effectively leveraging the multiple side information signals. This also calls for a method that avoids recovery performance degradation when the correlation among the prior information signals and the target signal decreases. As such, unlike prior studies [29, 30, 34], our method distributes weights across multiple side information signals. We propose to solve the problem in (11) such that at every iteration \( k \) we iterate over (i) a weight update step (i.e., updating \( W_j \) given \( x \)) and (ii) a data computation step (i.e., computing \( x \) given \( W_j \)).

**Updating \( W_j \) given \( x \) and \( z_j \).** Fixing \( x \) and \( z_j, j = 1, \ldots, J \), we determine the weights \( \{w_{ji}\}_j \) so as to improve the reconstruction of \( x \), where \( j \) indexes the side information signal and \( i = 1, \ldots, n \) iterates over the data elements. To normalize the contribution of each side information signal during the iterative process, we impose a constraint on the weights. We may have different strategies to update the weights \( \{w_{ji}\}_j \); in this work, use the constraint \( \sum_{j=0}^{J} W_j = I_n \), where \( I_n \) is the identity matrix with dimensions \( n \times n \). Consider the \( i \)-th element in \( x \) in Problem (11): we want to assign \( w_{ji} > 0 \) to the related term \( x_i - z_{ji} \), where \( z_{ji} \) is the \( i \)-th element in \( z_j \). We aim at assigning to each element \( x_i - z_{ji} \) a weight \( w_{ji} \) that is high when \( |x_i - z_{ji}| \) is low; in this way, we can favor \( x_i \) components that are closer to \( z_{ji} \). To ensure that zero-valued elements \( |x_i - z_{ji}| \) do not cause a breakdown of the method, we add a small parameter \( \epsilon > 0 \) and we set

\[
w_{ji} = \frac{\eta_i}{|x_i - z_{ji}| + \epsilon}, \tag{12}
\]
where \( \eta_i > 0 \). Under the constraint \( \sum_{j=0}^{J} W_j = I_n \), we obtain
\[
\eta_i = \left( \frac{1}{\sum_{j=0}^{J} |x_i - z_{ji}| + \epsilon} \right)^{-1}.
\] (13)
Using (13), we can rewrite each weight \( w_{ji} \) as
\[
w_{ji} = \frac{(|x_i - z_{ji}| + \epsilon)^{-1}}{\sum_{l=0}^{J} (|x_i - z_{li}| + \epsilon)^{-1}}.
\] (14)

**Computing \( x \) given \( W_j \).** Given \( W_j \), RAMSI computes \( x^{(k)} \) at iteration \( k \) using (3), where the proximal operator \( \Gamma_{\frac{1}{\lambda} g}(x_i) \) is computed as follows.

**Proposition 3.1.** The proximal operator \( \Gamma_{\frac{1}{\lambda} g}(x_i) \) in (4) for the problem of signal recovery with multiple side information, for which \( g(x) = \lambda \sum_{j=0}^{J} \| W_j (x - z_j) \|_1 \), is given by
\[
\Gamma_{\frac{1}{\lambda} g}(x_i) = \begin{cases} 
  x_i - \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)} & \text{if (16a)} \\
  z_{li} & \text{if (16b)}
\end{cases}
\] (15)
where
\[
z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)} < x_i < z_{(r+1)i},
\] (16a)
\[
z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r-1<j)} \leq x_i \leq z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)},
\] (16b)
and where, without loss of generality, we have assumed that \( -\infty = z_{(-1)i} \leq z_{0i} \leq z_{1i} \leq \ldots \leq z_{Ji} \leq z_{(J+1)i} = \infty \), and we have defined a boolean function
\[
b(r < j) = \begin{cases} 
  1, & \text{if } r < j \\
  0, & \text{otherwise}
\end{cases}
\] (17)

with \( r \in \{-1, \ldots, J\} \).

**Proof.** The proof is given in Appendix B.

We summarize RAMSI in Algorithm 1 (the Matlab code is provided in [55]), which is based on the fast iterative soft-thresholding algorithm (FISTA) algorithm [7]. The Stopping criterion (cf. Line 2 of Algorithm 1) can be a maximum iteration number \( k_{\text{max}} \), a relative variation of the objective function \( H(x) \) in (11), or a change of the number of nonzero components of the estimate \( x^{(k)} \). In this work, the relative variation of \( H(x) \) is chosen as stopping criterion.

**4. Bounds For Weighted \( n-\ell_1 \) Minimization**

We now establish measurement bounds for weighted \( n-\ell_1 \) minimization, serving as lower bounds for RAMSI.
Algorithm 1: The proposed RAMSI algorithm.

Input: $y$, $\Phi$, $z_1$, $z_2$, ..., $z_J$;

Output: $\hat{x}$;

1. Initialization : $W^{(1)}_0 = I$; $W^{(1)}_j = 0$ $\forall \ 1 \leq j \leq J$; $u^{(1)} = x^{(0)} = 0$; $L = L\nabla f$; $\lambda > 0$; $\epsilon > 0$; $t_1 = 1$; $k = 0$;

2. while Stopping criterion is false do
   3. $k = k + 1$;
   4. $\nabla f(u^{(k)}) = \Phi^T(\Phi u^{(k)} - y)$;
   5. $x^{(k)} = \Gamma \frac{1}{L} g(u^{(k)} - \frac{1}{L} \nabla f(u^{(k)}))$; where $\Gamma \frac{1}{L} g(.)$ is given by (15);

   // Updating weights.
   6. $w^{(k+1)}_{ji} = \frac{(|x_i^{(k)} - z_{ji}| + \epsilon)^{-1}}{\sum_{i=0}^{J} (|x_i^{(k)} - z_{ij}| + \epsilon)^{-1}}$;

   // Updating values for next iteration.
   7. $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$;
   8. $u^{(k+1)} = x^{(k)} + \frac{t_k - 1}{t_{k+1}} (x^{(k)} - x^{(k-1)})$;

9. end

10. return $x^{(k)}$;

4.1. Measurement Bound

4.1.1. Signal Traits

We begin our analysis by stating Definitions 4.1, 4.2, and 4.3, which do not limit the generality of our analysis but help us formalizing our bounds. Let $s_j$ denote the support of each difference vector $x - z_j$; namely, $\|x - z_j\|_0 = s_j$, where $j \in \{0, \ldots, J\}$ and $z_0 = 0$.

Definition 4.1. There are $0 \leq p < n$ indices for which the values in all $\{x - z_j\}_{j=0}^J$ vectors are nonzero and there are $n - q$ indices, with $0 < q \leq n$, for which the values in all $\{x - z_j\}_{j=0}^J$ vectors are equal to zero.

By rearranging the elements of all difference vectors in the same ordering, we can write the difference vectors as

$$x - z_0 = (x_1, \ldots, x_p, x_{p+1}, \ldots, x_q, 0, \ldots, 0),$$

$$x - z_1 = (x_1 - z_{11}, \ldots, x_p - z_{1p}, x_{p+1} - z_{1(p+1)}, \ldots, x_q - z_{1q}, 0, \ldots, 0),$$

$$\ldots$$

$$x - z_J = (x_1 - z_{J1}, \ldots, x_p - z_{Jp}, x_{p+1} - z_{J(p+1)}, \ldots, x_q - z_{Jq}, 0, \ldots, 0).$$

(18)
Consider the following example:
\[
\begin{align*}
  x &= (3, 0, 0, -2, 0, 0, 4, 0, 0, 0, 7), \\
  x - z_1 &= (0, 0, 3, 0, 0, 0, 2, 0, 0, -1, 0), \\
  x - z_2 &= (4, 0, 6, 0, 0, 3, -1, 0, 0, 0, 8).
\end{align*}
\]

In this example, \( n = 11, s_0 = 4, s_1 = 3, \) and \( s_2 = 5. \) Since at index \( i = 7 \) the elements of all \( \{x - z_j\}_{j=0}^2 \) vectors are nonzero, we have \( p = 1. \) Moreover, the elements of all vectors are zero at index \( i = 2, 5, 8, 9, \) hence \( q = 7. \) After rearranging the elements in the same ordering of the vectors as in (18), we obtain
\[
\begin{align*}
  x &= (4, 0, 0, -2, 0, 0, 7, 0, 0, 0, 0), \\
  x - z_1 &= (2, 0, 3, 0, -1, 0, 0, 0, 0, 0), \\
  x - z_2 &= (-1, 4, 6, 0, 3, 0, 8, 0, 0, 0, 0).
\end{align*}
\]

It is evident that \( p \leq \min\{s_j\} \) and \( q \geq \max\{s_j\}. \) As such, \( p \) and \( q \) can be seen as parameters that express the common part of the support and of the zero positions across the difference vectors. It can be noted that if \( q = 0, \) \( x = z_j = 0 \) and if \( p = n, \) \( \|x\|_0 = n, \) where \( x \) is not a sparse source as our consideration.

**Definition 4.2.** Following the rearrangement in (18), at any given position \( i \in \{p + 1, \ldots, q\}, \) we define \( d_i \) as the number of zero elements out of the \( J + 1 \) elements \( \{x_i - z_{ji}\}_{j=0}^J. \) Under Definition 4.1, we have that \( d_i \in \{1, \ldots, J\}. \)

In the example in (19), at \( i = 3 \) we have one zero element across the vectors, thus \( d_3 = 1. \) At \( i = 5 \) there are two zero elements, hence, \( d_5 = 2. \)

Using Definitions 4.1 and 4.2, we can express the total number of zero elements in (18) as
\[
(J + 1)(n - q) + \sum_{i=p+1}^{q} d_i = (J + 1)n - \sum_{j=0}^{J} s_j.
\]

**Definition 4.3.** Let \( i \) be a given position at the signal and multiple side information vectors. For any index \( i, \) without loss of generality, rearrange the \( J + 1 \) elements \( \{x_i - z_{ji}\}_{j=0}^J \) in order given by increasing values order into \( \{x_i - z_{ji}\}_{j=0}^J, \) i.e., \( -\infty = z_{(-1)i} \leq z_{0i} \leq z_{1i} \leq \ldots \leq z_{ji} \leq z_{(J+1)i} = \infty, \) where, for convenience, we introduced \( z_{(-1)i} \) and \( z_{(J+1)i} \) to denote \( -\infty \) and \( \infty, \) respectively. It can be noted that we rearrange for each individual index \( i. \) Under Definitions 4.1 and 4.2, we define \( r_i \in \{-1, \ldots, J\} \) such that \( x_i \in (z_{r_i i}, z_{(r_i+1)i}], \) or alias, \( \text{sign}(x_i - z_{ji}) = (-1)^{b(r_i < j')}, \) where \( b(r_i < j') \) is defined in (17).

At the example in (19), at index \( i = 1, \) \( x_1 = 4 \) and \( \{z_{j'1}\}_{j'=1}^{3} = \{-\infty, 0, 2, 5, \infty\}. \) Thus \( x_1 \in (z_{11}, z_{21}], \) i.e., \( r_1 = 1. \) For convenience, in the rest of the paper, we use notation \( j \) instead of \( j' \) denoting the reordered indices.
4.1.2. The Measurement Bound

Based on Definitions 4.1, 4.2, and 4.3, we shall derive a bound on the number of measurements required by weighted \(n-\ell_1\) minimization to successfully recover the target signal. Our generic bound is defined by Theorem 4.4, while a simpler but looser bound is described in Section 4.2.

**Theorem 4.4.** The number of measurements \(m_{n-\ell_1}\) required by weighted \(n-\ell_1\) minimization to recover the signal \(x\), given measurements \(y = \Phi x\) and \(J\) side information signals \(z_j\), is bounded as

\[
m_{n-\ell_1} \geq 2\bar{a}_{n-\ell_1} \log \frac{n}{\bar{s}_{n-\ell_1}} + \frac{7}{5} \bar{s}_{n-\ell_1} + \delta_{n-\ell_1} + 1,
\]

(21)

where \(\bar{a}_{n-\ell_1}\), \(\bar{s}_{n-\ell_1}\), and \(\delta_{n-\ell_1}\) are defined as

\[
\bar{a}_{n-\ell_1} = \sum_{i=1}^{p} a_i^2,
\]

(22a)

\[
\bar{s}_{n-\ell_1} = p + \sum_{i=p+1}^{q} (1 - c_i),
\]

(22b)

\[
\delta_{n-\ell_1} = (\kappa_{n-\ell_1} - 1)(\bar{s}_{n-\ell_1} - p),
\]

(22c)

wherein

\[
a_i = \sum_{j=0}^{J} w_{ji} (-1)^{b_{r_i} < j}, \quad c_i = d_i \left( \sum_{j=0}^{J} [x_i - z_{ji}] \right)^{-1} \epsilon \text{ with } \epsilon > 0, \text{ and}
\]

\[
\kappa_{n-\ell_1} = \frac{2 \cdot \min\{c_i\}}{\sqrt{\pi} \log(n/\bar{s}_{n-\ell_1}) [2 \cdot \min\{c_i\} - 1]}.
\]

(23)

**Proof.** The proof is given in Appendix C. \(\square\)

The code computing the quantities is provided in [55].

4.2. Further Analysis and Comparison with Known Bounds

We now relate the derived bound in Theorem 4.4 with the bounds of compressed sensing and compressed sensing with prior information, which are reported in Sec. 2.

**Corollary 4.4.1 (Relations with the known bounds).** The bound \(m_{n-\ell_1}\) for weighted \(n-\ell_1\) minimization in (21) becomes

(a) the \(\ell_1\)-minimization bound \(m_{\ell_1}\) in (5), when \(W_0 = I_n\) and \(W_j = 0\) for \(j \in \{1, \ldots, J\}\), that is,

\[
m_{n-\ell_1} \equiv m_{\ell_1} \geq 2s_0 \log \frac{n}{s_0} + \frac{7}{5} s_0 + 1,
\]

(24)

(b) the \(\ell_1-\ell_1\) minimization bound \(m_{\ell_1-\ell_1}\) in (8), when \(W_0 = W_1 = \frac{1}{2} I_n\) and \(W_j = 0\) for \(j \in \{2, \ldots, J\}\), that is,

\[
m_{n-\ell_1} \equiv m_{\ell_1-\ell_1} \geq 2\tilde{h} \log \frac{n}{s_{\ell_1-\ell_1}} + \frac{7}{5} s_{\ell_1-\ell_1} + 1,
\]

(25)

where \(\tilde{h}\) is given by (9b) and \(s_{\ell_1-\ell_1} = \frac{s_0 + s_1}{2}\).
Proof. The proof is given in Appendix C.

We now compare our bound for weighted $n$-\ell_1 minimization with the \ell_1-\ell_1 minimization [29, 30, 34] bound in (8). To this end, we derive a simplified bound in (26), which approximates our bound in (21). Furthermore, we introduce two simplicified bounds, which are looser bounds—one for each method—that are independent from the values of $x_j$, $z_j$.

**The simplified bound.** Our bound for weighted $n$-\ell_1 minimization in (21) becomes approximately

$$
\tilde{m}_{n,\ell_1} \geq 2\bar{a}_{n,\ell_1} \log \frac{n}{p} + \frac{7}{5} p + 1.
$$

(26)

Our approximation has the following reasoning: Firstly, Lemma Appendix D.1 in Appendix D proves that the $\delta_{n,\ell_1}$ term in (21) is negative; hence, the bound in (26) is looser. Secondly, since $\epsilon > 0$ is very small, we have $c_i \to 1^-$ [see Lemma Appendix D.1 for a quick explanation]. Consequently, $\bar{s}_{n,\ell_1} \approx p$ and $\delta_{n,\ell_1} \approx 0$ from (22b) and (22c), respectively; thereby, leading to our approximation. This simple bound is easier to evaluate compared to the bound in (21), as we only need to compute $\bar{a}_{n,\ell_1}$ and $p$. Furthermore, according to (22a) and Definition 4.1, we can write that $\bar{a}_{n,\ell_1} \leq p \leq \min\{s_j\}$, that is, $p$ in (26) is smaller than $\bar{s}_{0}$ in (5) and $\bar{s}_{\ell_1,\ell_1}$ in (25). Consequently, the simple bound in (26) for weighted $n$-\ell_1 minimization is sharper than the $\ell_1$ minimization and the $\ell_1$-\ell_1 minimization bounds, which means that the bound in (21) is even sharper. For the special case that $p = 0$, $\bar{a}_{n,\ell_1} = 0$ from (22a), $\bar{s}_{n,\ell_1} \approx p$, $\delta_{n,\ell_1} \approx 0$, consequently, from (21) $\tilde{m}_{n,\ell_1} \geq 1$.

**The simplified bounds with element-value independence.** Weighted $n$-\ell_1 minimization and $\ell_1$-\ell_1 minimization have looser bounds that are independent from the values of $x_j$, $z_j$, as given by

$$
\tilde{m}_{\ell_1,\ell_1} \geq 2p \log \frac{n}{\bar{s}_{\ell_1,\ell_1}} + \frac{7}{5} \bar{s}_{\ell_1,\ell_1} + 1,
$$

(27)

$$
\tilde{m}_{\ell_1,\ell_1} \geq 2p \log \frac{n}{\bar{s}_{\ell_1,\ell_1}} + \frac{7}{5} \bar{s}_{\ell_1,\ell_1} + 1,
$$

(28)

where $p$ is defined in Definition 4.1, $p = \min\{s_0, s_j\}$, and $\bar{s}_{\ell_1,\ell_1} = \frac{s_0 + s_j}{2}$. We obtain these bounds as follows. The bounds in (26) and (25) depend on the values of $x_j$, with $j \in \{1, \ldots, J\}$, via the quantities $\bar{a}_{\ell_1,\ell_1}$ and $\bar{h}$, respectively. From (22a) and (9b), we observe that $\bar{a}_{n,\ell_1} \leq p \leq \min\{s_j\}$ and $\bar{h} \leq \min\{s_0, s_1\}$. Hence, by replacing $\bar{a}_{n,\ell_1}$ and $\bar{h}$ with their maximum value leads to the looser bounds in (27) and (28).

The bounds in (27) and (28) reveal the advantage of using multiple side information signals in compressed sensing: It is evident that, by definition, $p \leq \rho$; hence, $\tilde{m}_{n,\ell_1} \leq \tilde{m}_{\ell_1,\ell_1}$ showing that the $n$-\ell_1 bound is better than the $\ell_1$-\ell_1 bound. Moreover, the higher the number of side information signals the smaller the bound is expected to become (because $p \leq \min\{s_j\}$).
Furthermore, according to (28), if the side information signal $z_1$ is not good enough, i.e., if $s_1 \gg s_0$, then $\hat{m}_{\ell_1-\ell_1} > m_{\ell_1}$, because of $\bar{s}_{\ell_1-\ell_1} \gg s_0$, thereby highlighting the limitations of the $\ell_1-\ell_1$ minimization method compared to the proposed weighted $n-\ell_1$ minimization approach.

Finally, we compare our bound to the existing bound of the weighted $\ell_1$ minimization with the prior support, $m_{w,\ell_1}$, in (6). This bound depends on the support $s_0$ and the support estimate’s error $s_e$ from prior information that reveals that $m_{w,\ell_1}$ in (6) is worse than $\hat{m}_{w,\ell_1}$ in (26). If we do not have good support for the estimate, i.e., high values of $s_e$, then $m_{w,\ell_1}$ in (6) is worse than $m_{\ell_1}$ in (5), which is illustrated in Sec. 5.

5. Experimental Results

5.1. Experimental Setup

We consider the reconstruction of a synthetic sparse signal $x$ given known prior signals $z_j$, $j = 1, 2, 3$. We generate $x$ from the i.i.d. zero-mean, unit-variance Gaussian distribution, with $n = 1000$ and $s_0 = 128$. Firstly, we consider the scenario where the side information signals $z_j$ are highly correlated with the signal $x$; in this case, $z_1$, $j = 1, 2, 3$, are generated such that they satisfy $s_j = \|x - z_j\|_0 = 64$ similar to the generation in [29, 30]. Moreover, a parameter is controlling the number of positions of nonzeros for which both $x$ and $x - z_j$ coincide. Let us denote this number by $\zeta_j$. For instance, if $\zeta_j = 51$, $x$ has 51 nonzero positions that coincide with 51 nonzero positions of $x - z_j$. This incurs a significant error between the source and the side information, given by $\|z_j - x\|_2/\|x\|_2 \approx 0.56$. We also generate prior information signals that are poorly correlated with $x$; in this case, we have $s_j = 256$ and $s_j = 352$. Furthermore, we set $\zeta_j \approx 128$, namely, 128 nonzero positions of $x$ coincide with 128 nonzero positions out of the total 256 or 352 nonzero positions of the side information signals, $z_j$, $j = 1, 2, 3$. This leads to very high errors, e.g., $\|z_j - x\|_2/\|x\|_2 \approx 1.12$ for $s_j = 256$, and the supports $s_j$ of $x - z_j$ are much higher than that of $x$. To constrain the number of cases, we set all $s_j$ equal.

5.2. Performance Evaluation

5.2.1. Signal Recovery Accuracy

We now assess the performance of RAMSI against state-of-the-art methods. Furthermore, we evaluate the bounds for weighted $n-\ell_1$ minimization [cf. (21)] against the bounds for classical CS [cf. (5)], $\ell_1-\ell_1$ minimization [cf. (8)], weighted-$\ell_1$ and Modified-CS [cf. (6)]. For a fixed number of measurements $m$, the probability of successful recovery $Pr(success)$ is the number of times $x$ is recovered as $\hat{x}$ with an error $\|\hat{x} - x\|_2/\|x\|_2 \leq 10^{-2}$, divided by the total number of 100 Monte-Carlo iterations (trials), where each trial considers different generated $x$, $z_1$, $z_2$, $z_3$, $\Phi$.

Let RAMSI-$J-\ell_1$ denote the RAMSI algorithm that uses $J$ side information signals, that is, one ($z_1$), two ($z_1$, $z_2$), or three ($z_1$, $z_2$, $z_3$), where we set $\epsilon = 10^{-5}$, $\lambda = 10^{-5}$. The existing FISTA
Figure 2: Successful probabilities of the original 1000-D $x$ vs. number of measurements $m$ for RAMSI using one, two, three side information signals. [7], FISTA-$\ell_1$-$\ell_1$ [30], Mod-CS [21], Reg-Mod-CS [22], and Weighted-$\ell_1$ [25] algorithms are used for comparison, where FISTA-$\ell_1$-$\ell_1$ denotes the $\ell_1$-$\ell_1$ method that uses only one side information ($z_1$).

The support estimate of $x$ that is considered as prior information by the Mod-CS, Reg-Mod-CS, and Weighted-$\ell_1$ algorithms is given by the support of $z_1$. Let $m_3, m_2, m_1, m_{\ell_1}, m_{\ell_1}, m_{\text{modCS}}$, and $m_{\text{FISTA}}$ denote the bounds of RAMSI-3-$\ell_1$, RAMSI-2-$\ell_1$, RAMSI-1-$\ell_1$, FISTA-$\ell_1$-$\ell_1$, Weighted-$\ell_1$, Mod-CS, and FISTA.

Fig. 2(a) depicts the $\Pr(\text{success})$ versus the number of measurements and Table 1 [Column 2] depicts the bounds for $s_j = 64$. The results show clearly that RAMSI-3-$\ell_1$ gives the sharpest bound and the highest recovery performance. For such good quality of side information, the accurate support...
Table 1: Measurement bounds for $s_0 = 128$ and different $s_j = 64, 256, 352$ values of $s_1 = s_2 = s_3 = s_j$

<table>
<thead>
<tr>
<th>Bound</th>
<th>$s_j = 64$</th>
<th>$s_j = 256$</th>
<th>$s_j = 352$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{3, \ell_1}$</td>
<td>208</td>
<td>414</td>
<td>414</td>
</tr>
<tr>
<td>$m_{2, \ell_1}$</td>
<td>229</td>
<td>451</td>
<td>451</td>
</tr>
<tr>
<td>$m_{1, \ell_1}$</td>
<td>272</td>
<td>520</td>
<td>519</td>
</tr>
<tr>
<td>$m_{\ell_1, \ell_1}$ [30]</td>
<td>259</td>
<td>485</td>
<td>523</td>
</tr>
<tr>
<td>$m_\omega \ell_1 \equiv m_{\text{modCS}}$ [21, 25]</td>
<td>253</td>
<td>911</td>
<td>NA</td>
</tr>
<tr>
<td>$m_{\ell_1}$ [7]</td>
<td>713</td>
<td>713</td>
<td>713</td>
</tr>
</tbody>
</table>

The estimate leads to that the bound $m_\omega \ell_1 \equiv m_{\text{modCS}}$ is better than $m_{1, \ell_1}$, $m_{\ell_1, \ell_1}$, and $m_{\ell_1}$ [see Column 2 in Table 1]. Furthermore, the performance of RAMSI-2-\ell_1 is higher than those of RAMSI-1-\ell_1 and FISTA-\ell_1-\ell_1. In this scenario, where the side information is of high quality, FISTA-\ell_1-\ell_1 outperforms our RAMSI-1-\ell_1 method. Furthermore, Reg-Mod-CS outperforms RAMSI-1-\ell_1, FISTA-\ell_1-\ell_1, Mod-CS, and Weighted-\ell_1. In addition, the weighted-\ell_1 method performs better than FISTA but worse than the remaining schemes. Hence, in this case, the use of equal weights—as in FISTA-\ell_1-\ell_1—leads to a higher performance than using adaptive weights with only one side information signal, as it is the case for RAMSI-1-\ell_1. This can be explained by comparing the $m_{1, \ell_1}$ and the $m_{\ell_1, \ell_1}$ bounds: It is clear that $\bar{a}_{1, \ell_1}$ [see (22a)] is greater than $\bar{h}$ [see (4b)]. By combining this observation with the small value of $s_1$—due to the good-quality side information signal $z_1$—, which in turn results in a small $\bar{s}_{\ell_1, \ell_1}$ value, explains why the bound as well as the recovery performance of FISTA-\ell_1-\ell_1 are better than those of RAMSI-1-\ell_1 [see magenta and black lines in Fig. 2(a)]. We can conclude that by exploiting multiple side information signals we can obtain the best performance and when dealing with only one good-quality side information signal we may choose to reconstruct using equal weights.

Figs. 2(b) and 2(c) present the reconstruction performance versus the number of measurements and Table 1 [Column 3 and 4] present the bounds when the side information signals are less correlated with the signal of interest, that is, $s_j = 256$ and $s_j = 352$. In this scenario, all RAMSI configurations outperform the FISTA, FISTA-\ell_1-\ell_1, Weighted-\ell_1, Reg-Mod-CS, and Mod-CS methods. The performance of RAMSI-1-\ell_1 is better than those of FISTA-\ell_1-\ell_1, Weighted-\ell_1, Reg-Mod-CS, and Mod-CS. In Table 1 [Column 3], we also see that the bound $m_\omega \ell_1 \equiv m_{\text{modCS}}$ in (6) is the worst since the support estimate $s_e$ in (6) is high. In Table 1 [Column 4] the bound for Modified-CS is not available since the support estimate’s error $s_e > s_0$ is out of the consideration of the bound $m_\omega \ell_1$ in (6). Interestingly, in Fig. 2(c), we observe that the accuracy of FISTA-\ell_1-\ell_1 is worse than that of FISTA, i.e., the side information $z_1$ does not help, however, RAMSI-1-\ell_1 still outperforms FISTA. These results highlight the drawback of the $\ell_1$-\ell_1 minimization method when the side information is of poor quality. Despite
using poor side information signals, all RAMSI versions achieve better results than FISTA due to the proposed re-weighted $n$-$\ell_1$ minimization algorithm. We observe that the performance of RAMSI-2-$\ell_1$ and RAMSI-3-$\ell_1$ is slightly worse than that of RAMSI-1-$\ell_1$. These small penalties are due to the poor quality of the side information signals, which has an impact on the iterative update process.

5.2.2. Side Information Quality-Dependence Analysis

We now assess the impact of the side information quality on the number of measurements required to successfully reconstruct the target signal. We consider different values for $s_1 = s_2 = s_3$, ranging from 20 to 400 (where $s_j$ is the support size of the difference vector $x - z_j$). The signal $x$ has $s_0 = 128$ and it is generated as in Sec. 5.1. For a fixed value of $s_1 = s_2 = s_3$, we report the number of measurements required by RAMSI-3-$\ell_1$, RAMSI-2-$\ell_1$, RAMSI-1-$\ell_1$, FISTA-$\ell_1$-$\ell_1$, FISTA to achieve a probability of successful recovery bounded as $\text{Pr}(\text{success}) \geq 0.98$.

Fig. 3 shows that the recovery performance is clearly improved when leveraging side information; and the more signals are considered the better the performance. However, when the support size of $x - z_j$ meets that of the target signal, the side information does not improve the performance anymore. When $s_j > s_0$, the performance and the measurement bound of FISTA-$\ell_1$-$\ell_1$ are increasing with the value of $s_1$. Specifically, when $s_1 > 315$, the performance of FISTA-$\ell_1$-$\ell_1$ is worse than that of FISTA, thereby illustrating the limitations of $\ell_1$-$\ell_1$ minimization when the side information is of poor quality. As shown in Fig. 3, the performance of RAMSI—both in terms of the bounds and the practical results—is robust against poor-quality side information. The bounds are sharper when the number of side information signals increases and remain approximately constant when $s_j$ increases after a threshold (indicating increasing-poor side information quality). When $s_j > 300$, the number of measurements of RAMSI-3-$\ell_1$ and RAMSI-2-$\ell_1$ are slightly worse than those of RAMSI-1-$\ell_1$, approaching the number of measurements of FISTA [this behavior was also observed in Fig. 2(c)]. To address this issue, we can adaptively select the best among the RAMSI configurations. For instance, when one side information is available, we can choose to use equal weights in case of good side information. Furthermore, we can ensure that RAMSI’s performance is not worse than FISTA by weighing dominantly on the signal rather than on the side information signals. To do so, we need to make an estimate of the quality of the side information signals, which is left as a topic for future research.

6. Conclusion

We proposed the RAMSI algorithm and established measurement bounds for the problem of signal recovery under a weighted $n$-$\ell_1$ minimization framework. RAMSI incorporates multiple side information signals in the problem of sparse signal recovery and iteratively weights them so as to optimize the
reconstruction performance. The bounds confirm the advantage of RAMSI in utilizing multiple side information signals to significantly reduce the number of measurements and to deal with variations in the quality of the side information. We experimentally assessed the bounds and the performance of RAMSI against state-of-the-art methods using numerical sparse signals. The results showed that our bounds are sharper than existing bounds and that RAMSI outperforms the conventional CS method, the recent $\ell_1$-$\ell_1$ minimization method as well as the Modified-CS, regularized Modified-CS, and weighted $\ell_1$ minimization methods.

Appendix A. Background on Measurement Bounds

We summarise some key definitions and conditions in convex optimization and linear inverse problems [11, 12], which are used in the derivation of the measurement bounds for the proposed weighted $n$-$\ell_1$ minimization approach.

Appendix A.0.1. Convex Cone

A convex cone $C \subset \mathbb{R}^n$ is a convex set that satisfies $C = \tau C$, $\forall \tau \geq 0$ [12]. For the cone $C \subset \mathbb{R}^n$, a polar cone $C^\circ$ is the set of outward normals of $C$, defined by

$$C^\circ := \{ u \in \mathbb{R}^n : u^T x \leq 0, \ \forall x \in C \}. \quad (A.1)$$

A descent cone [12, Definition 2.7] $\mathcal{D}(g, x)$, alias tangent cone [11], of a convex function $g := \mathbb{R}^n \rightarrow \mathbb{R}$ at a point $x \in \mathbb{R}^n$—at which $g$ is not increasing—is defined as

$$\mathcal{D}(g, x) := \bigcup_{\tau \geq 0} \{ y \in \mathbb{R}^n : g(x + \tau y) \leq g(x) \}. \quad (A.2)$$
where \( \bigcup \) denotes the union operator.

**Appendix A.0.2. Gaussian Width**

The Gaussian width [11] is a summary parameter for convex cones; it is used to measure the aperture of a convex cone. For a convex cone \( C \subset \mathbb{R}^n \), considering a subset \( C \cap \mathbb{S}^{n-1} \) where \( \mathbb{S}^{n-1} \subset \mathbb{R}^n \) is a unit sphere, the Gaussian width [11, Definition 3.1] is defined as

\[
\omega(C) := \mathbb{E}_g \left[ \sup_{u \in C \cap \mathbb{S}^{n-1}} g^T u \right],
\]

where \( g \sim N(0, I_n) \) is a vector of \( n \) independent, zero-mean, and unit-variance Gaussian random variables and \( \mathbb{E}_g [\cdot] \) denotes the expectation with respect to \( g \). The Gaussian width [11, Proposition 3.6] can further be bounded as

\[
\omega(C) \leq \mathbb{E}_g [\text{dist}(g, C^\circ)],
\]

where \( \text{dist}(g, C^\circ) \) denotes the Euclidean distance of \( g \) with respect to the set \( C^\circ \), which is in turn defined as

\[
\text{dist}(g, C^\circ) := \min_u \{ \| g - u \|_2 : u \in C^\circ \}.
\]

Recently, a summary parameter called the statistical dimension \( \delta(C) \) of the cone \( C \) [12] is used to estimate the convex cone [12, Theorem 4.3]. The statistical dimension is expressed in terms of the polar cone \( C^\circ \) as [12, Proposition 3.1]

\[
\delta(C) := \mathbb{E}_g [\text{dist}^2(g, C^\circ)].
\]

From (A.4) and (A.6), we can derive the following inequality

\[
\omega^2(C) \leq \delta(C),
\]

which gives a convenient bound for the Gaussian width that will be used in our computations.

**Appendix A.0.3. Measurement Condition**

An optimality condition [11, Proposition 2.1], [12, Fact 2.8] for linear inverse problems states that \( x_0 \) is the unique solution of (2) if and only if

\[
\mathcal{D}(g, x_0) \cap \text{null}(\Phi) = \{0\},
\]

where \( \text{null}(\Phi) := \{ x \in \mathbb{R}^n : \Phi x = 0 \} \) is the null space of \( \Phi \). We consider the number of measurements \( m \) required to successfully recover a given signal \( x_0 \in \mathbb{R}^n \). Corollary 3.3 in [11] states that, given a measurement vector \( y = \Phi x_0 \), \( x_0 \) is the unique solution of (2) with probability at least \( 1 - \exp(-\frac{1}{2} (\sqrt{m} - \omega(\mathcal{D}(g, x_0)))^2) \) provided that \( m \geq \omega^2(\mathcal{D}(g, x_0)) + 1 \). Using (A.7), we can write this condition as

\[
m \geq \delta(\mathcal{D}(g, x_0)) + 1.
\]
Appendix A.0.4. Bound on the Measurement Condition

Using (A.6), the statistical dimension $\delta(D(g, x))$ of a descent cone $D(g, x)$ is

$$\delta(D(g, x)) = \mathbb{E}_g \left[ \text{dist}^2(g, D(g, x))^\circ \right],$$

(A.10)

where $D(g, x)^\circ$ is the polar cone of $D(g, x)$ defined in (A.1). Let us consider that the subdifferential $\partial g$ [56] of a convex function $g$ at a point $x \in \mathbb{R}^n$ is given by $\partial g(x) := \{u \in \mathbb{R}^n : g(y) \geq g(x) + u^T(y - x) \text{ for all } y \in \mathbb{R}^n \}$. From (A.10) and [12, Proposition 4.1], we obtain a lower bound on $\delta(D(g, x))$ as

$$\delta(D(g, x)) \geq \mathbb{E}_{\tau \geq 0} \left[ \min \text{dist}^2(g, \tau \cdot \partial g(x)) \right] \leq \min_{\tau \geq 0} \mathbb{E}_g \left[ \text{dist}^2(g, \tau \cdot \partial g(x)) \right].$$

(A.11)

In short, we conclude the following proposition.

**Proposition Appendix A.1** (Measurement bound for a convex norm function). In order to obtain the measurement bound for the recovery condition, $m \geq U_g + 1$, we calculate the quantity $U_g$ given a convex norm function $g := \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$U_g = \min_{\tau \geq 0} \mathbb{E}_g \left[ \text{dist}^2(g, \tau \cdot \partial g(x)) \right].$$

(A.12)

Appendix B. Proof of Proposition 3.1

Replacing $g(x) = \lambda \sum_{j=0}^J \|W_j(x - z_j)\|_1$ in the definition of the proximal operator in (4), we obtain

$$\Gamma_{\lambda g}(x) = \arg \min_{v \in \mathbb{R}} \left\{ \frac{\lambda}{2} \sum_{j=0}^J \|W_j(v - z_j)\|_1 + \frac{1}{2} \|v - x\|^2 \right\}.$$  

(B.1)

Both terms in (B.1) are separable in $v$ and thus, we can minimize each element $v_i$ of $v$ individually as

$$\Gamma_{\lambda g}(x_i) = \arg \min_{v_i \in \mathbb{R}} \left\{ h(v_i) = \frac{\lambda}{2} \sum_{j=0}^J w_{ji} |v_i - z_{ji}| + \frac{1}{2} (v_i - x_i)^2 \right\}. $$

(B.2)

To solve (B.2), we need to derive $\frac{\partial h(v_i)}{\partial v_i}$. Recall from Definition 4.3 that, without loss of generality, we assume that $-\infty = z_{i(-1)} \leq z_{0i} \leq z_{1i} \leq \cdots \leq z_{(J+1)i} = \infty$. When $v_i \in (z_{ri}, z_{(r+1)i})$ with $r \in \{-1, \ldots, J\}$, $h(v_i)$ is differentiable and $\frac{\partial h(v_i)}{\partial v_i}$ is calculated as

$$\frac{\partial h(v_i)}{\partial v_i} = \frac{\lambda}{L} \sum_{j=0}^J w_{ji} \text{sign}(v_i - z_{ji}) + (v_i - x_i),$$

(B.3)

where $\text{sign}(\cdot)$ is the sign function. Using the definition of the $b(\cdot)$ function in (17), we have that $\text{sign}(v_i - z_{ji}) = (-1)^{b(r<j)}$ and thus, we can rewrite (B.3) as

$$\frac{\partial h(v_i)}{\partial v_i} = \frac{\lambda}{L} \sum_{j=0}^J w_{ji} (-1)^{b(r<j)} + (v_i - x_i).$$

(B.4)
By setting $\frac{\partial h(v_i)}{\partial v_i} = 0$, we obtain:

$$v_i = x_i - \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)}. \quad (B.5)$$

Since $v_i \in (z_{ri}, z_{(r+1)i})$, via (B.5) we have that

$$z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)} < x_i < z_{(r+1)i} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)}. \quad (B.6)$$

In the following Lemma Appendix B.1, we prove that, in the remaining range value of $x_i$, namely, in the case that

$$z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)} \leq x_i \leq z_{ri} + \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}(-1)^{b(r<j)}, \quad (B.7)$$

the minimum of $h(v_i)$ is obtained when $v_i = z_{ri}$.

**Lemma Appendix B.1.** Given that $x_i$ is bounded as in (B.7), the function $h(v_i)$ defined in (B.2) is minimized when $v_i = z_{ri}$.

**Proof.** Let us express $h(v_i)$ as

$$h(v_i) = \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} |v_i - z_{ri}| - \lambda \sum_{j=0}^{J} w_{ji} |z_{ri} - z_{ri}| + \frac{1}{2} |v_i - z_{ri}|^2 - \lambda \sum_{j=0}^{J} w_{ji} |x_i - z_{ri}| + \frac{1}{2} |x_i - z_{ri}|^2. \quad (B.8)$$

Applying the inequality $|a - b| \geq |a| - |b|$, with $a, b \in \mathbb{R}$, on the first summation term and expanding the second term, we obtain:

$$h(v_i) \geq \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} |v_i - z_{ri}| - \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} |z_{ri} - z_{ri}| + \frac{1}{2} (v_i - z_{ri})^2 - \lambda \sum_{j=0}^{J} w_{ji} |x_i - z_{ri}| + \frac{1}{2} (x_i - z_{ri})^2. \quad (B.9)$$

Using the basic inequality $-(v_i - z_{ri}) \cdot (x_i - z_{ri}) \geq -|v_i - z_{ri}| \cdot |x_i - z_{ri}|$, we can express (B.9) as

$$h(v_i) \geq (v_i - z_{ri}) \cdot \left( \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} |v_i - z_{ri}| \cdot |x_i - z_{ri}| + \frac{1}{2} (v_i - z_{ri})^2 - \lambda \sum_{j=0}^{J} w_{ji} |x_i - z_{ri}| + \frac{1}{2} (x_i - z_{ri})^2 \right). \quad (B.10)$$

At this point, via (B.7) we obtain

$$- \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} \leq x_i - z_{ri} \leq \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} \Rightarrow |x_i - z_{ri}| \leq \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji}. \quad (B.11)$$

We now observe that the part including $v_i$ in the right hand side of (B.10) can be written as

$$|v_i - z_{ri}| \left( \frac{\lambda}{L} \sum_{j=0}^{J} w_{ji} - |x_i - z_{ri}| \right) + \frac{1}{2} (v_i - z_{ri})^2. \quad (B.12)$$

Taking into account (B.11), the expression in (B.12) and, in turn $h(v_i)$, are minimized when $v_i = z_{ri}$. 

\[\square\]
In summary, via (B.5) and Lemma Appendix B.1, we obtain
\[
\Gamma \frac{1}{L} g(x_i) = \begin{cases} 
  x_i - \lambda \sum_{j=0}^{J} w_{ji} (-1)^{b(r<j)}, & \text{if } (B.6) \\
  z_{ri}, & \text{if } (B.7). 
\end{cases}
\quad (B.13)
\]

Appendix C. Proof of Theorem 4.4 and Corollary 4.4.1

We begin by stating some important results that help us proving Theorem 4.4 and Corollary 4.4.1.

Recall that the probability density of the normal distribution $N(0, 1)$ with zero-mean and unit variance $\psi(x)$ is given by
\[
\psi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. 
\quad (C.1)
\]

Our formulations consider the following inequality, which is stated in [30], that is
\[
\frac{1 - x^{-1}}{\sqrt{\pi \log(x)}} \leq \frac{1}{\sqrt{2\pi}} \leq \frac{2}{5}, 
\quad (C.2)
\]
for all $x > 1$. Moreover, adhering to the formulations in [30], we use the following expressions in our derivations:
\[
A(x) := \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} (v-x)^2 e^{-v^2/2} dv, 
\quad (C.3a)
\]
\[
B(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} (v-x)^2 e^{-v^2/2} dv, 
\quad (C.3b)
\]
for which we have that
\[
A(0) = B(0) = 1/2. 
\quad (C.4)
\]

When $x \neq 0$, the following inequalities have been derived in [30]:
\[
A(x) \leq \begin{cases} 
  \psi(x)/x, & x > 0 \\
  x^2 + 1, & x < 0 
\end{cases} 
\quad (C.5a)
\]
\[
B(x) \leq \begin{cases} 
  -\psi(x)/x, & x < 0 \\
  x^2 + 1, & x > 0. 
\end{cases} 
\quad (C.5b)
\]

Lemma Appendix C.1. Given $x \in (0, 1]$ and $\tau > 0$ for $\psi(x)$ given in (C.1), we have
\[
\frac{\psi(\tau x)}{\tau x} \leq \frac{1}{\sqrt{2\pi}} \frac{1 - x^2}{\tau x} + \frac{\psi(\tau)}{\tau}. 
\quad (C.6)
\]

Proof. Using (C.1) the left hand side of (C.6) becomes
\[
\frac{\psi(\tau x)}{\tau x} = \frac{1}{\sqrt{2\pi}} e^{-\tau^2 x^2/2}. 
\quad (C.7)
\]
Applying Bernoulli’s inequality on $e^{-\tau^2 x^2/2}$ leads to
\[
e^{-\tau^2 x^2/2} = \left(1 + (e^{-\tau^2/2} - 1)\right)^{x^2} \leq 1 + x^2(e^{-\tau^2/2} - 1), 
\quad (C.8)
\]
where $0 < x \leq 1$ and $(e^{-\tau^2/2} - 1) > -1$ given that $\tau > 0$. Combining (C.7) and (C.8) leads to the proof. □
Proof of Theorem 4.4. We derive the bound based on Proposition Appendix A.1 by firstly computing the subdifferential \( \partial g(x) \) and then the distance between the standard normal vector \( g \) to \( \partial g(x) \). Under Definitions 4.2 and 4.3, let us recall that there are \( d_i \) consecutive zero elements at any \( i \in \{p+1,\ldots,q\} \), namely, \( \{x_i - z_j\}_{j=1}^{l_i} = 0 \), with \( l_i \in \{0,\ldots,J\} \) being an auxiliary index indicating the start of the zero positions. Under Definitions 4.1, 4.2, 4.3, and the weights in (14), the elements of the vectors \( u \) in the subdifferential \( \partial g(x) \) are expressed as

\[
\begin{align*}
  u_i &= a_i, \quad i \in \{1,\ldots,p\}, \\
  u_i &= [b_i - c_i, b_i + c_i], \quad i \in \{p+1,\ldots,q\}, \\
  u_i &= [-\sum_{j=0}^{J} w_{ji}, \sum_{j=0}^{J} w_{ji}] = [-1,1], \quad i \in \{q+1,\ldots,n\},
\end{align*}
\]

where

\[
\begin{align*}
  a_i &= \sum_{j=0}^{J} w_{ji} (-1)^{b(i,j)}, \\
  b_i &= \sum_{j \in \{l_i,\ldots,l_i+d_i-1\}} w_{ji} (-1)^{b(i,j)}, \\
  c_i &= \sum_{j=l_i}^{J} w_{ji} = \frac{d_i}{e}.
\end{align*}
\]

Using (A.5) and (C.9), we can compute the distance from the standard normal vector \( g \) to the subdifferential \( \partial g(x) \) as

\[
\text{dist}^2(g, \tau : \partial g(x)) = \sum_{i=1}^{p} (g_i - \tau a_i)^2 + \sum_{i=p+1}^{q} \left( \frac{\beta^2}{2} (g_i - \tau (b_i + c_i)) + \beta^2 (-g_i + \tau (b_i - c_i)) \right) + \sum_{i=q+1}^{n} \beta^2 (|g_i| - \tau),
\] (C.11)

where \( \beta(a) := \max \{a, 0\} \) returns the maximum value between \( a \in \mathbb{R} \) and 0. Taking the expectation of (C.11) with respect to \( g \) delivers

\[
\begin{align*}
  \mathbb{E}_g [\text{dist}^2(g, \tau : \partial g(x))] &= p + \tau^2 \sum_{i=1}^{p} a_i^2 + \frac{1}{\sqrt{2\pi}} \sum_{i=p+1}^{q} \int_{(b_i+c_i)}^{\infty} (v - \tau (b_i + c_i))^2 e^{-v^2/2} dv \\
  &\quad + \frac{1}{\sqrt{2\pi}} \sum_{i=p+1}^{q} \int_{(b_i-c_i)}^{-\infty} (v - \tau (b_i - c_i))^2 e^{-v^2/2} dv + \sqrt{\frac{2}{\pi}} \sum_{i=q+1}^{n} \int_{\tau}^{\infty} (v - \tau)^2 e^{-v^2/2} dv.
\end{align*}
\] (C.12)

Replacing the expressions in (C.3a), (C.3b) in (C.12) gives

\[
\begin{align*}
  \mathbb{E}_g [\text{dist}^2(g, \tau : \partial g(x))] &= p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i=p+1}^{q} A \left( \tau (b_i + c_i) \right) + \sum_{i=p+1}^{q} B \left( \tau (b_i - c_i) \right) + 2 \sum_{i=q+1}^{n} A(\tau).
\end{align*}
\] (C.13)

At this point, it is worth emphasizing the advantage of the adaptive weights [see (14)] in the proposed method, the values of which depend on the correlation of the side information with the
target signal. Focusing on a given index \( i \in \{p + 1, \ldots, q \} \), let us observe the weight contribution in the expressions of \( b_i \) and \( c_i \), defined in (C.10b) and (C.10c), respectively. The weights \( w_{ji} \) in \( c_i \) are considerably higher than those in \( b_i \), as in the former case the side information signals \( z_{ji} \) are equal to the source \( x_i \). Moreover, we recall that a small positive parameter \( \epsilon \) is introduced in the denominator of the weights so as to avoid division by zero when \( x_i = z_{ji} \). The \( \epsilon \) parameter can ensure that \( c_i \) in (C.10c) is always greater than \( |b_i| \) in (C.10b). Hence, we always have \( b_i + c_i > 0 \) and \( b_i - c_i < 0 \). With these observations we conclude that the arguments of \( \mathcal{A}(\cdot) \) and \( \mathcal{B}(\cdot) \) in (C.13) are respectively positive and negative. Applying inequality (C.5a) for \( x > 0 \) on the expression of \( \mathcal{A}(\cdot) \) as well as inequality (C.5b) for \( x < 0 \) on the expression of \( \mathcal{B}(\cdot) \), we obtain the following bound for the \( U_g \) quantity [which is defined in (A.12)]:

\[
U_{g_n-t_1} \leq \min_{\tau \geq 0} \left\{ p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i=p+1}^{q} \psi(\tau(b_i + c_i)) \frac{\tau(b_i + c_i)}{\tau(b_i + c_i)} + \psi(\tau(c_i - b_i)) \frac{\tau(c_i - b_i)}{\tau(c_i - b_i)} + 2 \sum_{i=q+1}^{n} \psi(\tau) \right\},
\]

(C.14)

where \( \psi(\cdot) \) is the zero-mean, unit-variance normal distribution defined in (C.1). Applying (C.6) on the second sum in (C.14) gives

\[
U_{g_n-t_1} \leq \min_{\tau \geq 0} \left\{ p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i=p+1}^{q} \frac{(1 - (b_i + c_i)^2)}{\tau^2} \frac{1}{\tau^2} \left( \frac{1}{c_i - b_i} - 1 \right) + \left( n - q \right) + \sum_{i=p+1}^{q} c_i \right\} 2 \frac{\psi(\tau)}{\tau} \}
\]

(C.15)

\[
\Rightarrow U_{g_n-t_1} \leq \min_{\tau \geq 0} \left\{ p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i=p+1}^{q} \frac{1}{\tau^2} \left( \frac{1}{c_i - b_i} - 1 \right) + \left( n - q \right) + \sum_{i=p+1}^{q} c_i \right\} 2 \frac{\psi(\tau)}{\tau} \}
\]

(C.16)

From the definitions of \( b_i \) and \( c_i \) in (C.10b) and (C.10c), respectively, we have

\[
b_i = \sum_{j \in \{j_1, \ldots, j_{d_i}-1\}} w_{ji} = 1 - c_i,
\]

(C.17)

where we used the constraint \( \sum_{j=1}^{d_i} w_{ji} = 1 \). Hence, the second summation term in (C.16) is bounded as

\[
\sum_{i=p+1}^{q} \frac{1}{\tau^2} \left( \frac{1}{c_i - b_i} - 1 \right) \leq \sum_{i=p+1}^{q} \frac{4 \sqrt{2\pi \tau}}{2 \pi} \frac{c_i}{2 \pi} \left( \frac{1}{c_i} - 1 \right) \leq \frac{4 \cdot \min\{c_i\}}{\sqrt{2\pi \tau} \left( \min\{c_i\} - 1 \right)} \sum_{i=p+1}^{q} (1 - c_i),
\]

(C.18)

(C.19)

where (C.18) is obtained by using that \( b_i \leq 1 - c_i \) and (C.19) holds from the fact that the term \( \frac{c_i}{\sqrt{2\pi \tau} \left( \min\{c_i\} - 1 \right)} \) is maximized when \( c_i \) is minimized (recall that \( c_i > 0 \)). For simplicity, let us denote

\[
\bar{a}_{n-t_1} = \sum_{i=1}^{p} a_i^2,
\]

(C.20a)

\[
\kappa_{n-t_1} = \frac{4 \cdot \min\{c_i\}}{\sqrt{2\pi \tau} \left( \min\{c_i\} - 1 \right)},
\]

(C.20b)

\[
\bar{s}_{n-t_1} = q - \sum_{i=p+1}^{q} c_i = p + \sum_{i=p+1}^{q} (1 - c_i).
\]

(C.20c)
Substituting the quantities of (C.20a), (C.20b), (C.20c), and using inequality (C.19) in (C.16) gives

\[
U_{g_{n-\ell_1}} \leq \min_{\tau \geq 0} \left\{ p + \bar{a}_{n-\ell_1} \tau^2 + \kappa_{n-\ell_1} (\bar{s}_{n-\ell_1} - p) + (n - \bar{s}_{n-\ell_1}) \frac{2\psi(\tau)}{\tau} \right\},
\]

which using the definition of \(\psi(\cdot)\) can be written as

\[
U_{g_{n-\ell_1}} \leq \min_{\tau \geq 0} \left\{ \bar{a}_{n-\ell_1} \tau^2 + (n - \bar{s}_{n-\ell_1}) \frac{2 \sqrt{2/\pi \tau} e^{-\frac{\tau^2}{2\pi}}}{\tau} + \bar{s}_{n-\ell_1} + (\kappa_{n-\ell_1} - 1)(\bar{s}_{n-\ell_1} - p) \right\}. \tag{C.21}
\]

To derive a bound as a function of the source signal \(x\), we need to select a parameter \(\tau > 0\). Setting \(\tau = \sqrt{2 \log(n/\bar{s}_{n-\ell_1})}\) yields

\[
U_{g_{n-\ell_1}} \leq 2\bar{a}_{n-\ell_1} \log \frac{n}{\bar{s}_{n-\ell_1}} + \frac{\bar{s}_{n-\ell_1}(1 - \bar{s}_{n-\ell_1}/n)}{\sqrt{\pi \log(n/\bar{s}_{n-\ell_1})}} + \bar{s}_{n-\ell_1} + \delta_{n-\ell_1}. \tag{C.22}
\]

where

\[
\delta_{n-\ell_1} = (\kappa_{n-\ell_1} - 1)(\bar{s}_{n-\ell_1} - p), \tag{C.23}
\]

and where we have replaced the selected value of \(\tau\) in (C.20b), thereby obtaining the \(\kappa_{n-\ell_1}\) definition reported in (23).

Applying inequality (C.2) on the second term of the right hand side of (C.22) gives

\[
U_{g_{n-\ell_1}} \leq 2\bar{a}_{n-\ell_1} \log \frac{n}{\bar{s}_{n-\ell_1}} + \frac{\bar{s}_{n-\ell_1}(1 - \bar{s}_{n-\ell_1}/n)}{\sqrt{2\pi \tau}} + \bar{s}_{n-\ell_1} + \delta_{n-\ell_1}. \tag{C.24}
\]

Bearing in mind that \(m_{n-\ell_1} \geq U_{g_{n-\ell_1}} + 1\) and by combining (C.24) with (C.20), (C.10), and (C.23) leads to the proof.\(\Box\)

**Proof of Corollary 4.4.1.** We start with Relation (a). Under the conditions that \(W_j = I_{\ell_1}\) and \(W_j = 0\) for \(j \in \{1, \ldots, J\}\), we have that \(s_{n-\ell_1} = p = s_0\) and \(a_{n-\ell_1} = p = s_0\), where we used the definitions in (C.20a), (C.20c). Consequently, from (C.23), \(\delta_{n-\ell_1} = 0\). Replacing these values of \(a_{n-\ell_1}, s_{n-\ell_1}, \delta_{n-\ell_1}\) in our bound, defined in (21), leads to the \(\ell_1\) minimization bound in (5).

To reach Relation (b), given that \(W_0 = W_1 = \frac{1}{2}I_{\ell_1}\) and \(W_j = 0\) for \(j \in \{2, \ldots, J\}\), let us first denote two subsets, \(\Gamma_1\) and \(\Gamma_2\), as

\[
\Gamma_1 := \left\{ i \in \{p + 1, \ldots, q\} : b_i + c_i = 1, b_i - c_i = 0 \right\}. \tag{C.25a}
\]

\[
\Gamma_2 := \left\{ i \in \{p + 1, \ldots, q\} : b_i + c_i = 0, b_i - c_i = -1 \right\}. \tag{C.25b}
\]

Via the definitions of \(b_i\) and \(c_i\)—see (C.10b) and (C.10c), respectively—and since \(i \in \{p + 1, \ldots, q\}\), we observe that \(i \in \Gamma_1\) or \(i \in \Gamma_2\).

Replacing (C.25a) and (C.25b) in (C.13) leads to

\[
\mathbb{E}_x[\text{dist}^2(g, \tau \cdot \vartheta(g(x))) = p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i \in \Gamma_1} A(\tau) + \sum_{i \in \Gamma_1} B(0) + \sum_{i \in \Gamma_2} A(0) + \sum_{i \in \Gamma_2} B(-\tau) + 2 \sum_{i=q+1}^{n} A(\tau). \tag{C.26}
\]
By combining (C.4), (C.5a), and (C.5b) with (C.26), we obtain the following bound for the $U_g$ quantity [defined in (A.12)]:

$$U_{g_{\ell_1}, \epsilon_1} \leq \min_{\tau \geq 0} \left\{ p + \tau^2 \sum_{i=1}^{p} a_i^2 + \sum_{i=p+1}^{q} \frac{1}{2} + \sum_{i=q+1}^{n} \frac{\psi(\tau)}{\tau} + 2 \sum_{i=q+1}^{n} \frac{\psi(\tau)}{\tau} \right\}. \tag{C.27}$$

which can further be elaborated to

$$U_{g_{\ell_1}, \epsilon_1} \leq \min_{\tau \geq 0} \left\{ \tau^2 \sum_{i=1}^{p} a_i^2 + \frac{1}{2}(p + q) + (2n - (p + q)) \frac{e^{-\tau^2}}{\sqrt{2\pi\tau}} \right\}. \tag{C.28}$$

In the $\ell_1$-$\ell_1$ minimization case, there is a single side information signal and no weights; that is, $d_i = 1$ and $J = 1$ in (20) under Definitions 4.1 and 4.2; thus, we have $p + q = s_0 + s_1$. Combining this result with (9b) gives $\sum_{i=1}^{n} a_i^2 = \bar{h}$.

Let us now denote $\bar{s}_{\ell_1, \epsilon_1} = \frac{s_0 + s_1}{2}$ and set $\tau = \sqrt{2\log(n/\bar{s}_{\ell_1, \epsilon_1})}$; thus, we have

$$U_{g_{\ell_1}, \epsilon_1} \leq 2\bar{h} \log \frac{n}{\bar{s}_{\ell_1, \epsilon_1}} + \bar{s}_{\ell_1, \epsilon_1} + \frac{\bar{s}_{\ell_1, \epsilon_1}}{n} \left( \frac{1 - \bar{s}_{\ell_1, \epsilon_1}}{\bar{s}_{\ell_1, \epsilon_1}} \right) \sqrt{2\pi \log \left( \frac{n}{\bar{s}_{\ell_1, \epsilon_1}} \right)} \tag{C.29}$$

Applying (C.2) on the third term of the right hand side of (C.29) gives

$$U_{g_{\ell_1}, \epsilon_1} \leq 2\bar{h} \log \frac{n}{\bar{s}_{\ell_1, \epsilon_1}} + \frac{7}{5} \bar{s}_{\ell_1, \epsilon_1}. \tag{C.30}$$

Finally, we obtain the $\ell_1$-$\ell_1$ minimization bound [29, 30, 34] in (8) as

$$m_{\ell_1, \epsilon_1} \geq 2\bar{h} \log \frac{n}{\bar{s}_{\ell_1, \epsilon_1}} + \frac{7}{5} \bar{s}_{\ell_1, \epsilon_1} + 1, \tag{C.31}$$

where $\bar{s}_{\ell_1, \epsilon_1} = \frac{s_0 + s_1}{2} = s_0 + \frac{\epsilon}{2}$, with $\epsilon$ defined in (9a).

\section*{Appendix D. Supporting Lemmas}

\textbf{Lemma Appendix D.1.} The value of $\delta_{n, \ell_1}$ in the definition of the bound for weighted $n$-$\ell_1$ minimization—given in (21)—is negative.

\textbf{Proof.} Recall the definition of $\delta_{n, \ell_1}$ given in (C.23). By the definition of $\bar{s}_{n, \ell_1}$ in (C.20c), it is clear that $(\bar{s}_{n, \ell_1} - \bar{y}) > 0$; this is because $c_i < 1$ [see (C.10c)]. Hence, the sign of $\delta_{n, \ell_1}$ depends on the term $(\bar{s}_{n, \ell_1} - 1)$. From (C.20b), it is clear that $\kappa_{n, \ell_1}$ depends on $c_i$, which is defined as

$$c_i = d_i \left( d_i + \sum_{j \notin \{1, \ldots, d_i + 1\}} \frac{\epsilon}{|x_i - x_j| + \epsilon} \right)^{-1}. \tag{D.1}$$

As $\epsilon > 0$ and very small, we can say that $c_i \to 1^-$. As a result, $\kappa_{n, \ell_1}$ is approximately given by

$$\kappa_{n, \ell_1} \approx \frac{4}{\sqrt{2\pi}} \approx \frac{2}{\sqrt{\pi \log(n/\bar{s}_{n, \ell_1})}}, \tag{D.2}$$

26
where in the proof of Theorem 4.4 we have set that $\tau = \sqrt{2 \log(n/\bar{s}_n-\ell_1)}$. We observe that $\kappa_{n-\ell_1} < 1$ if $\frac{\bar{s}_n-\ell_1}{n} < 0.28$, where $\bar{s}_n-\ell_1 \approx p \leq \min\{s_j\}$ [see (C.20c) and Definition 4.1]. In Lemma Appendix D.2, we prove that when $\frac{p}{n} < 0.23$ then $\hat{m}_{n-\ell_1}$ in (27) is less than the source dimension, $n$. Otherwise, the required number of measurements $\hat{m}_{n-\ell_1}$ is higher than $n$, signifying the failure of the algorithm\(^2\). Hence, we get that $(\kappa_{n-\ell_1} - 1) < 0$, thereby proving that $\delta_{n-\ell_1} < 0$.

**Lemma Appendix D.2.** Given a sparse signal $x \in \mathbb{R}^n$ with $\|x\|_0 = s_0$ and side information signals $z_j \in \mathbb{R}^n$ with $\|x - z_j\|_0 = s_j$, $\forall j \in \{1, \ldots, J\}$, the number of measurements required for weighted $n-\ell_1$ minimization to recover $x$—given in (27)—is less than the source dimension $n$ when $\frac{p}{n} < 0.23$.

**Proof.** Let us assume that the number of measurements in the bound of (27) satisfies the condition $\hat{m}_{n-\ell_1} < n$, namely, we have

$$2p \log \frac{n}{p} + \frac{7}{5}p < n. \quad (D.3)$$

By substituting $\gamma = \frac{n}{p}$, $\gamma \in \mathbb{R}$, in (D.3) we get

$$2 \log \gamma + \frac{7}{5} < \gamma. \quad (D.4)$$

By setting $f(\gamma) = \gamma - 2 \log \gamma - \frac{7}{5}$ and using calculus, we can show that $f(\gamma) > 0$ when $\gamma > 4.33$, or else, $\frac{p}{n} < 0.23$.

---

**References**


\(^2\)Indeed, when the source signal is not sparse enough and the side information signals are not highly correlated with the source, the recovery algorithm will have a low performance.


[55] The matlab codes for the ramsi algorithm and the measurement bounds for weighted \( n-\ell_1 \) minimization. [Online]. Available: https://github.com/huynhlvd/ramsi